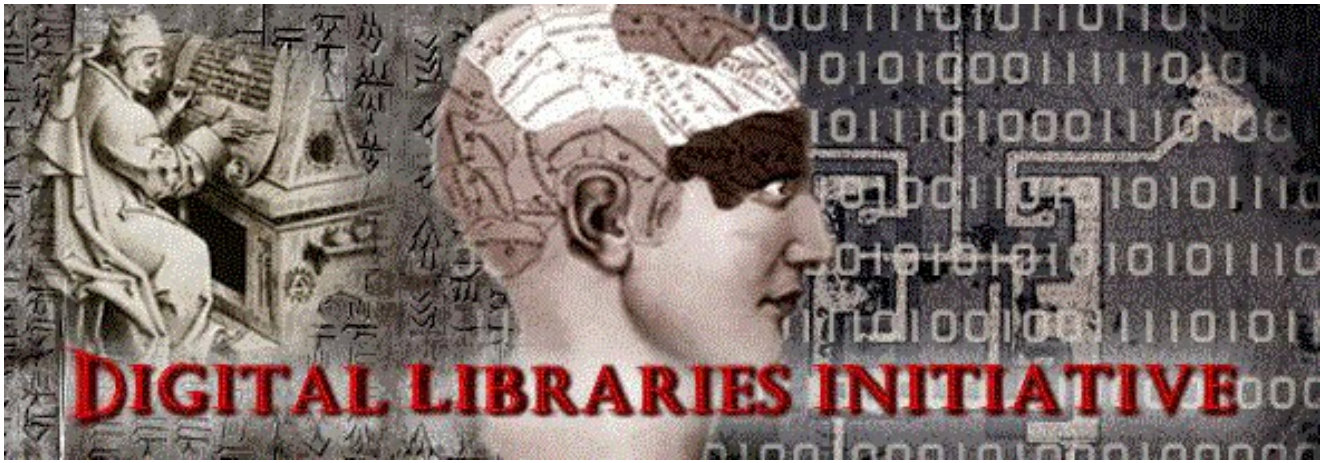
# DIGITAL LIBRARIES INITIATIVE
# An Interagency Program of Research and Applications




***www.dli2.nsf.gov***

Stephen M. Griffin
National Science Foundation

# Digital Libraries Initiative (DLI) Program Profile (Phase 1)

- Six university-led projects; similar project model for each
- $24M total over four years, ending fall 1998.
  (~$1M/year per project)
- Each project required to:
  - Carry out fundamental research
  - Create a large testbed
  - Work with partners
  - Acquire substantial cost sharing (1:1 is norm)
  - Demonstrate leadership for larger community
- Cooperative agreements, not grants
- All-Project meetings every six months
- D-Lib Magazine (DARPA sponsored)

# Digital Libraries Initiative (DLI) Phase 1 Projects

| Project/Research Focus | *research goal* |
|---|---|
| **Carnegie Mellon University: Digital Video Libraries**<br>• speech, image and natural language technologies integration | ⇨ ***full-content search and retrieval of video segments*** |
| **Univ of Michigan: Intelligent Agent Architectures**<br>• software agents**;** resource federation; artificial service market economies; educational impact | ⇨ ***new DL cross-disciplinary capabilities, intellectual perspectives and linkages*** |
| **Stanford Univ: Uniform Access**<br>• interoperability; protocols & standards; distributed object architectures; interface design for distributed information retreival | ⇨ ***general access, extensibility for heterogeneous distributed resources*** |

# Digital Libraries Initiative (DLI) Phase 1 Projects

| Project/Research Focus | *research goal* |
|---|---|
| **Univ of California, Santa Barbara: Geographic Information Systems**<br>• spatially-indexed data; content-based retrieval; image-compression; metadata | ⇨ ***resources for geosciences research and education communities*** |
| **Univ of Illinois: Intelligent Search and the Net**<br>• large-scale information retrieval across knowledge domains; semantic search; SGML; user/usage studies | ⇨ ***semantic retrieval across the net; alternatives for publishers of scientific journals*** |
| **Univ of California, Berkeley: Media Integration and Access**<br>• new models of “documents”; natural language processing; content-based image retrieval; innovative interface design | ⇨ ***new models and services for multi-media information management in a networked world*** |

# DLI Collaboration and Partnering

**-- DLI Lead Institutions --**

Carnegie Mellon University of California, Berkeley University of Illinois
Stanford University University of Michigan Univ of California, Santa Barbara

***Flow of Resources, Technologies, Knowledge, Intellectual Products***

**Computer & Communications Companies**
Digital Equipment Corp
Xerox Corp
Xerox PARC
Intel Corp
Apple Corporation
Bellcore
Eastman Kodak Co
IBM
Lockheed
Interconnect Tech Corp
Enterprise Integration (EIT)
Bellcore
Interval
Microsoft Corp
Bell Atlantic Network Services
AT&T
Hewlett Packard
United Technologies
Softquad
BRS/Dataware
Spyglass
Hitachi

**Publishers/Content Providers**
Elsevier Science Group
Encyclopedia Britannica
McGraw-Hill Publishers
Dialog Information Services
O'Reilly
WAIS Inc
QED Communications
John Wiley & Sons
U.S. News & World Report
M&T Publishing
Tribune Company

**Professional Societies**
American Math Society
UMI
(AMA)
ACM
IEEE
American Institute of Aeronautics and Astronautics (AIAA)
American Physical Society
American Institute of Physics
NCGIA
Association of Research

**Other Universities**
SUNY Buffalo
Univ of Maine
Univ of Arizona
Open University, U.K.
Univ of Wisconsin
Univ of Colorado
MIT
Cornell Univ

**Libraries**
Project Site Univ Libs
USGS Library
Library of Congress
California State Library
Sonoma County Library
St. Louis Public Library
New York Public Libs

**International Orgs**
ERCIM

**Primary & Secondary Schools**
Project-local comm schools
Fairfax County Public Schools
Winchester-Thurston School
Ann Arbor Public Schools
Stuyvesant High School, NYC
Shasta County Ofc of Edu

**Government Agencies and Labs**
DMA/CIO
U S Navy
USGS
NASA/ARC
Res Agcy of California
San Diego Assn of Govts

**Other/Non-Profits**
CNRI
Environmental Systems Res In
Mellon Foundation
Kellogg Foundation
Getty Foundation

# Digital Libraries Initiative – Phase 2

*www.dli2.nsf.gov*

## Sponsoring Agencies and Partners

**NSF (CISE, SBE, EHR/DUE)**
**DARPA**
**National Library of Medicine**
**Library of Congress**
**National Endowment for the Humanities**
**NASA**
**FBI**

**National Archives**
**Smithsonian**
**Institute for Museum and Library Services**

# *Digital Libraries Initiative - Phase 2*

**Core Sponsors: NSF, DARPA, NLM, LoC, NASA, NEH**

**~$8-10 million/yr for 4-5 years (beginning FY98)**

- **sponsor a full-spectrum of activities: fundamental research, content & collections development, domain applications, testbeds, operational environments, new resources for education and preserving America's cultural heritage**
- **address topics over entire DL lifecycle: information creation, dissemination, access, use, preservation, impact, contexts**
- **implement a modular, open program structure: add new sponsors, performers, projects at any time**

**Program Goals:**

**new DL research, technologies and applications to advance the use of distributed,**

**networked information of all types around the nation and the world**

# Comparison of DLI with DLI - Ph

## DLI

**research**
broad, technology-centered
**testbeds**
for technology research
**content/collections**
donated to projects
**infrastructure**
limited testbed development
**context**
primarily user evaluation

1994

*Information: http://dli.grainger.uiuc.edu/dli/national.htm*

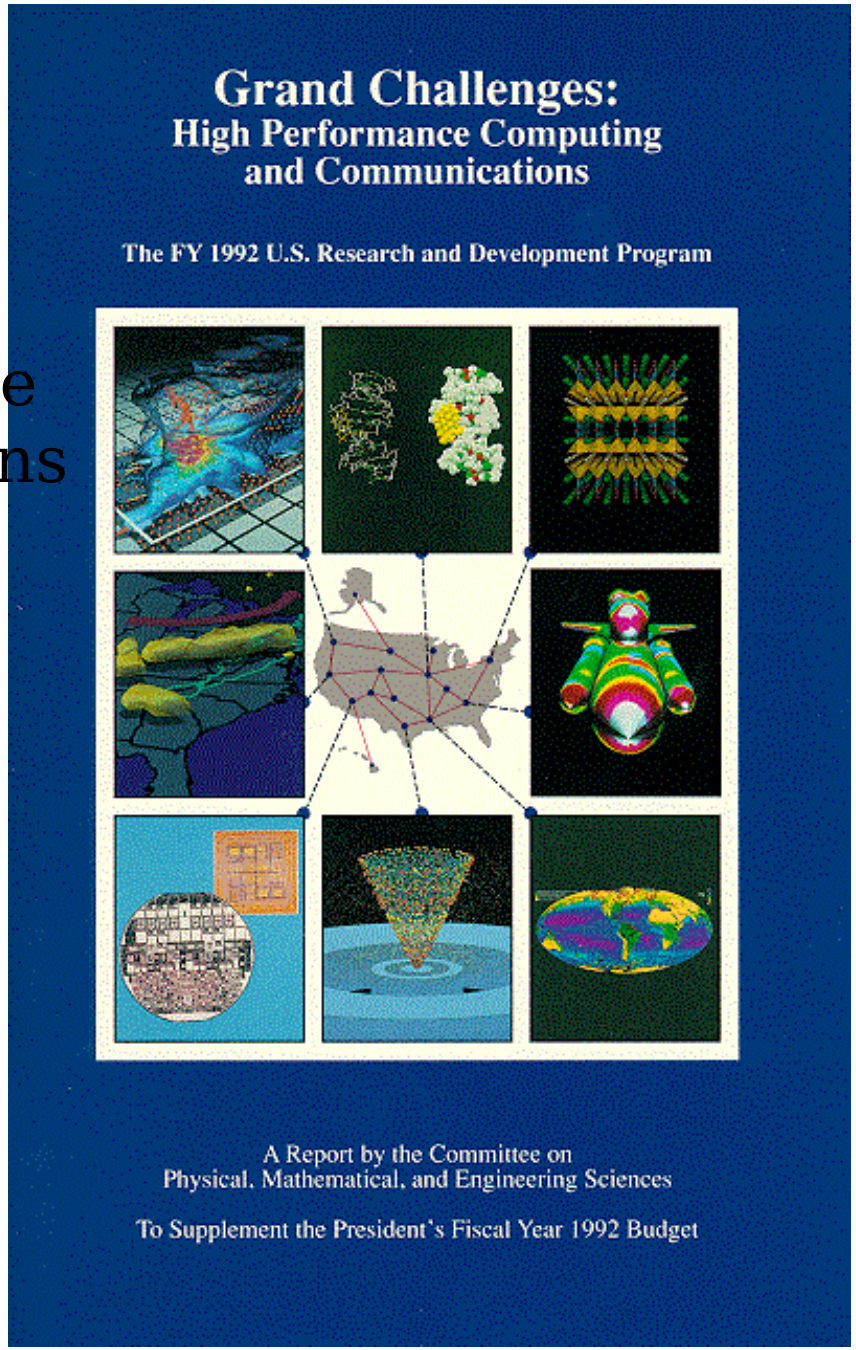
## DLI - Phase 2

**research**
refined technical scope; extend to new areas and dimensions in the DL information lifecycle
**testbeds**
for DL research with added emphasis on interoperability & technology integration
**content/collections**
increased emphasis on content, collections development and management
**infrastructure**
operational DLs with collections of value to domain and other "communities" of users
**context**
understanding DLs in domain

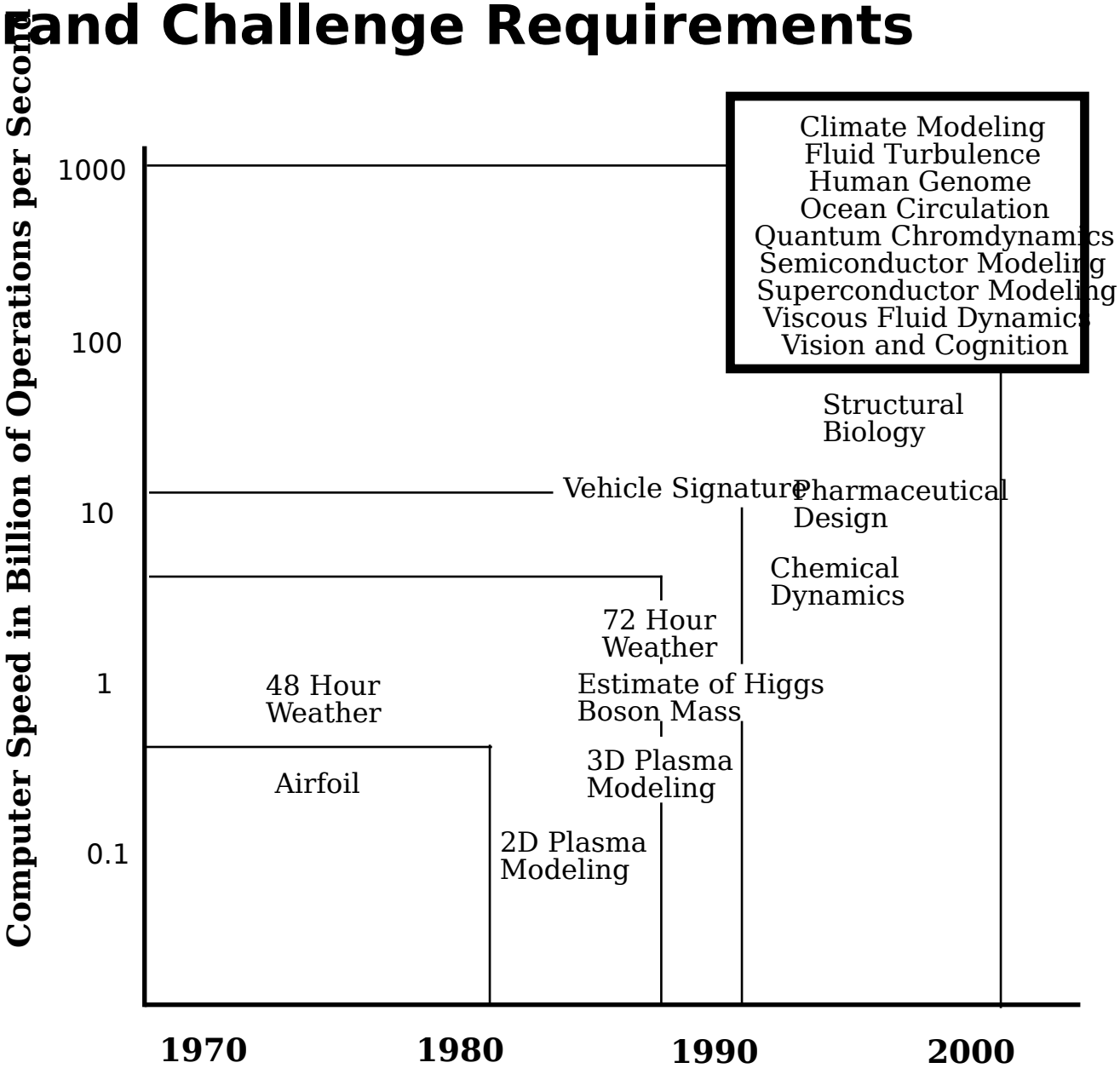
1998

The Federal High Performance Computing and Communications Program

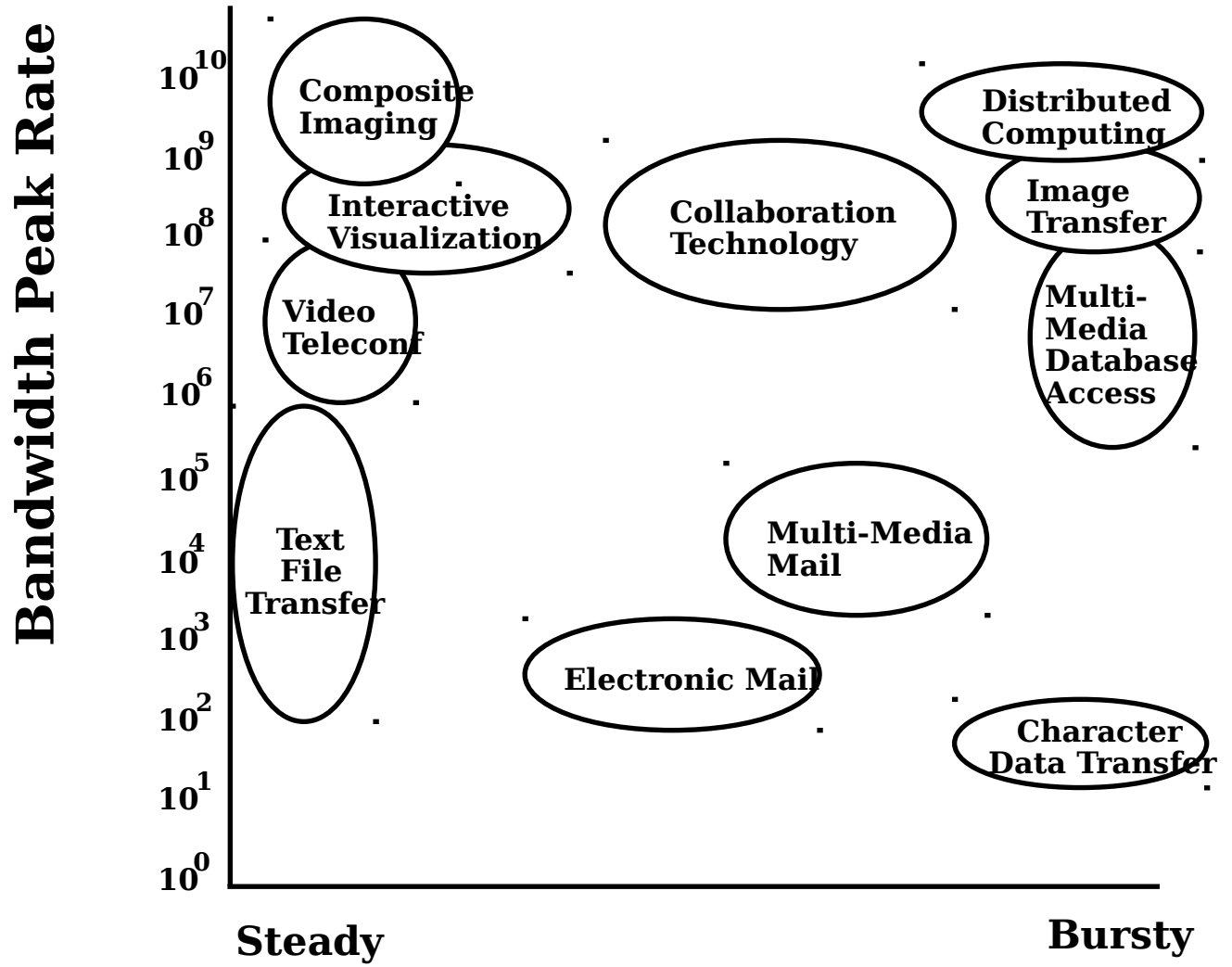
1992-1996


Grand Challenges:
High Performance Computing and Communications
The FY 1992 U.S. Research and Development Program
A Report by the Committee on
Physical, Mathematical, and Engineering Sciences
To Supplement the President's Fiscal Year 1992 Budget

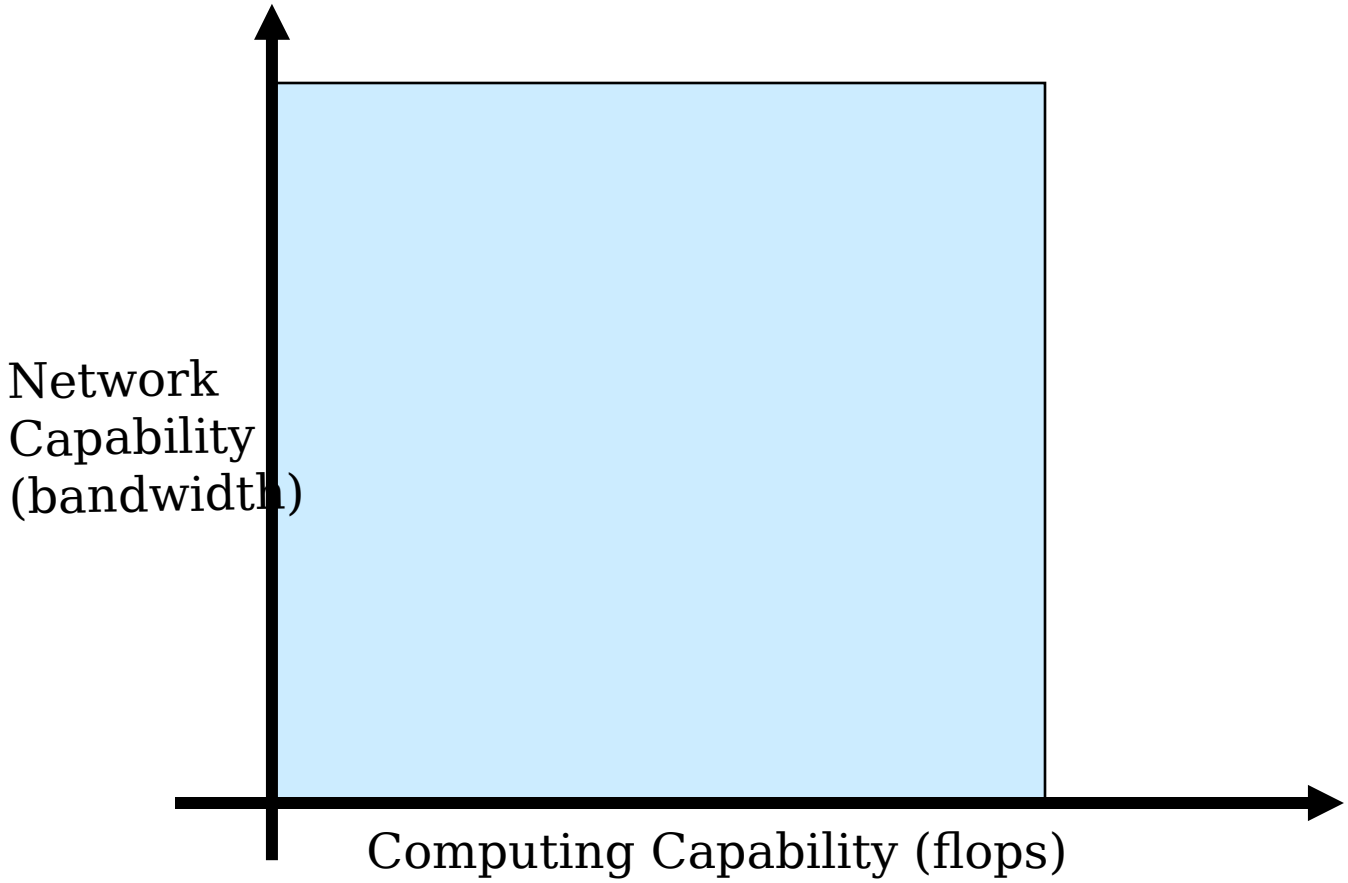
# Grand Challenge Requirements



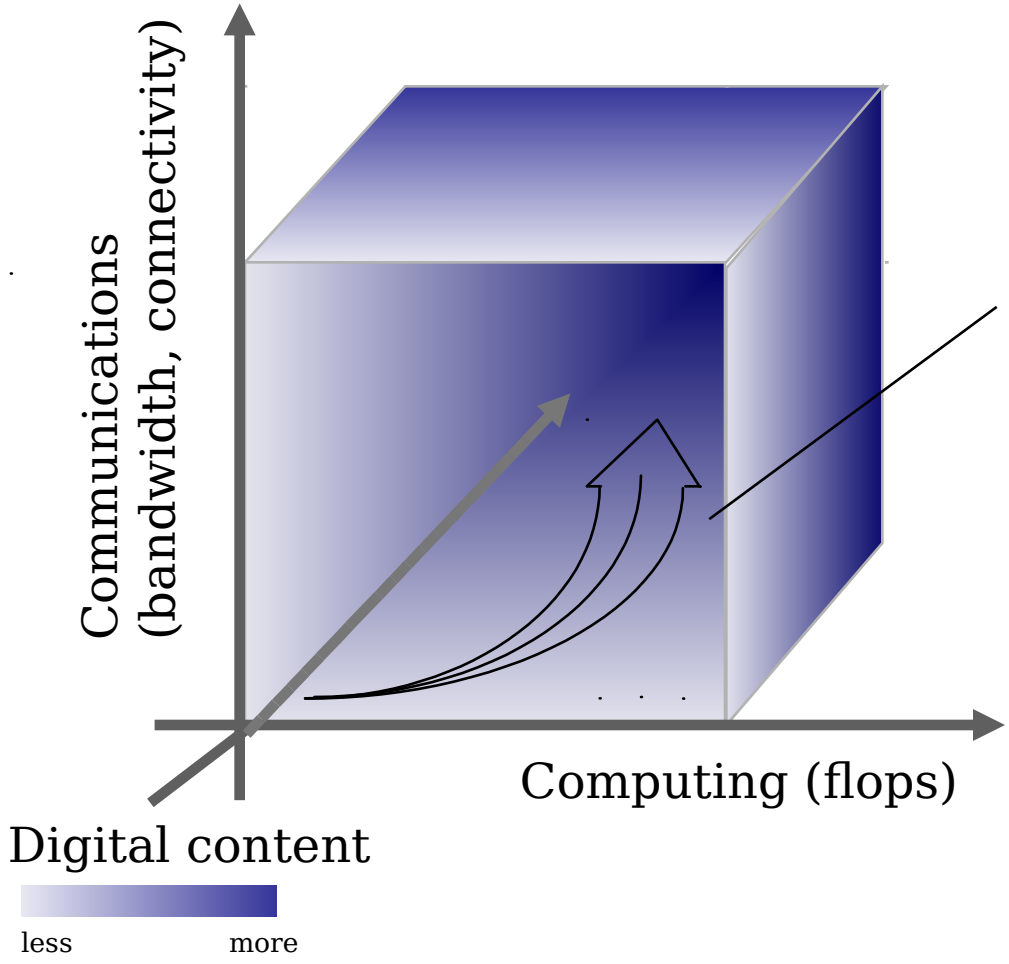
# NREN Applications
## by Bandwidth and Traffic Characteristics



# *Two Dimensional Thinking of Early 1990s…*

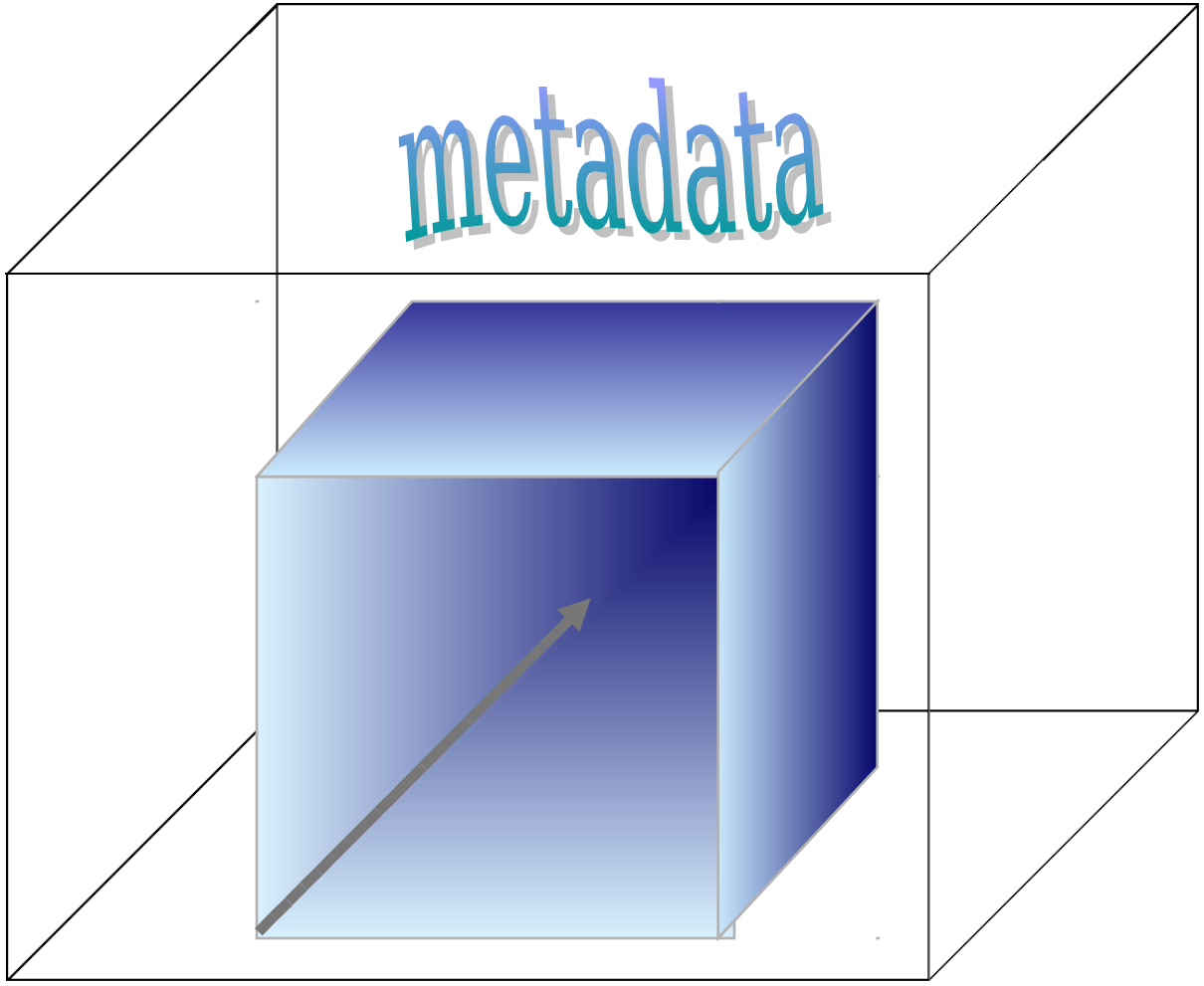


# *Three Dimensional Thinking of mid-90s*

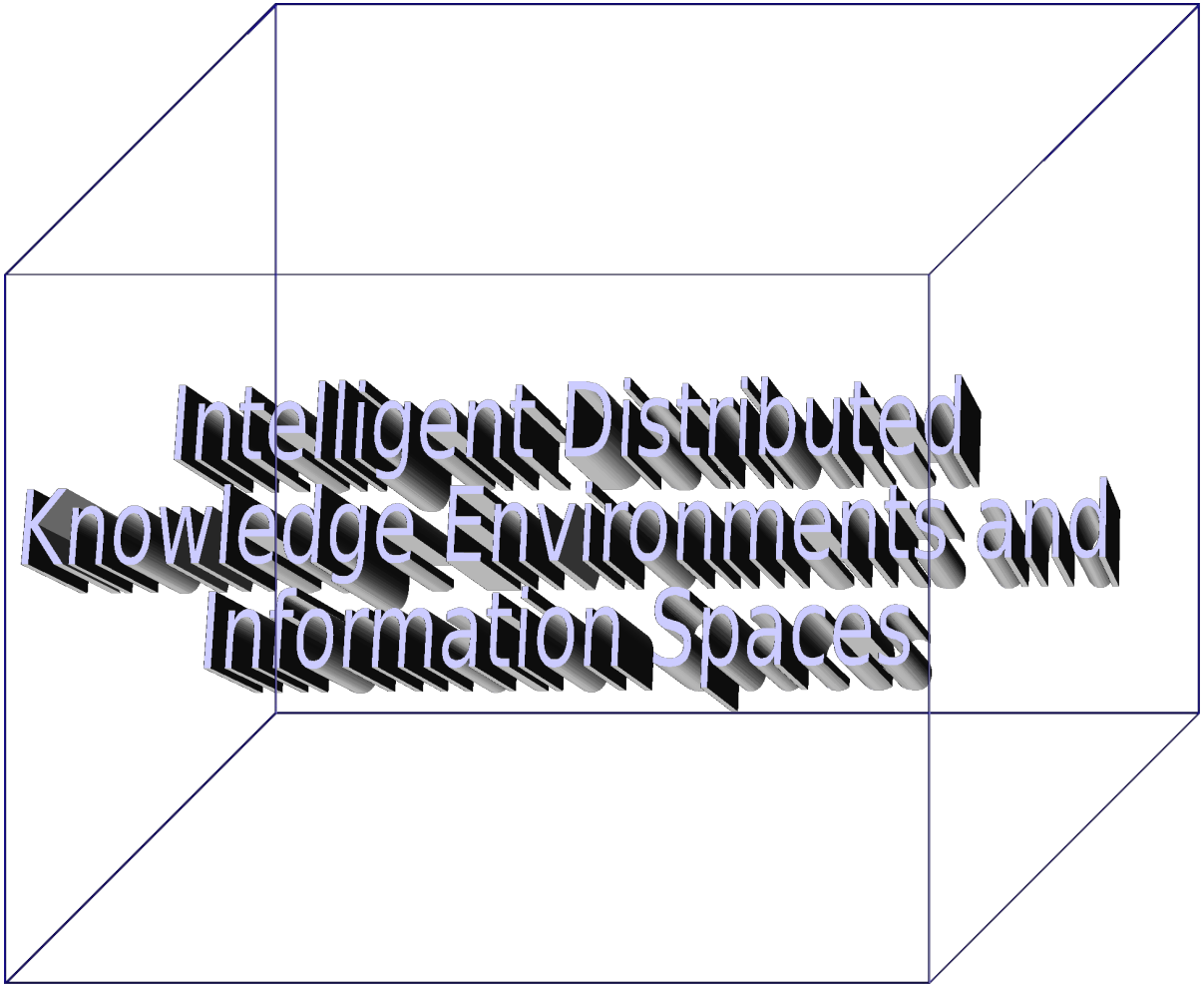



Digital Libraries technology trajectory: *intellectual access to globally distributed information*

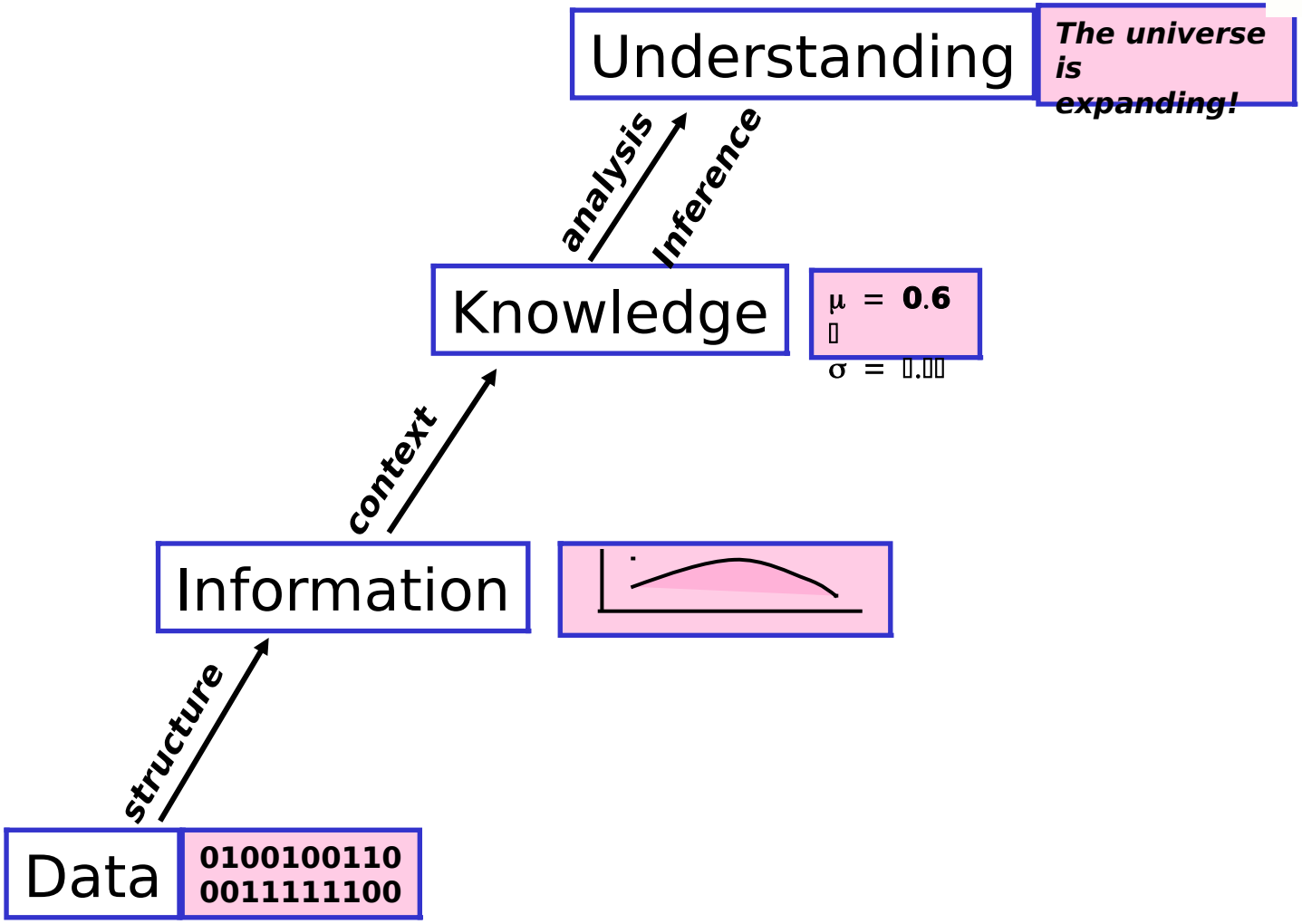
# Today: Emphasis on Context and Structure



# *Next: Advanced Functional Capabilities*



# *Evolution of Understanding in a Distributed Knowledge Environment*



# *Leibnitz and Newton: then and now*

## 1600s

1684: Leibnitz publishes paper on calculus
1693: Newton publishes paper on calculus
$10^1$ copies

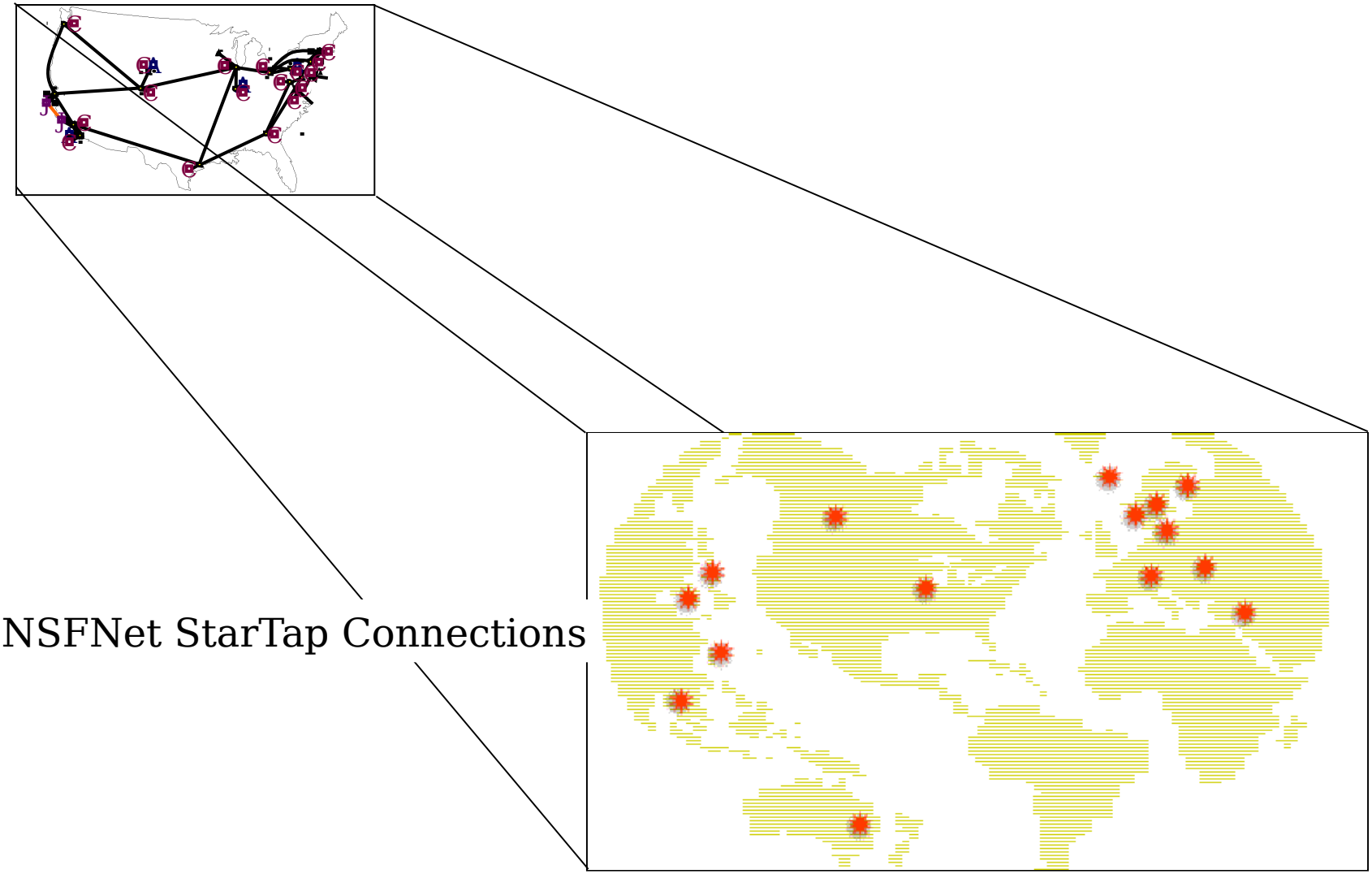
Newton receives credit ...

## Today...

May 31, 1999 15:24 GMT: Leibnitz' paper on pre-print server
May 31, 1999 15:33 GMT: Newton's paper on pre-print server
$10^n$ downloads

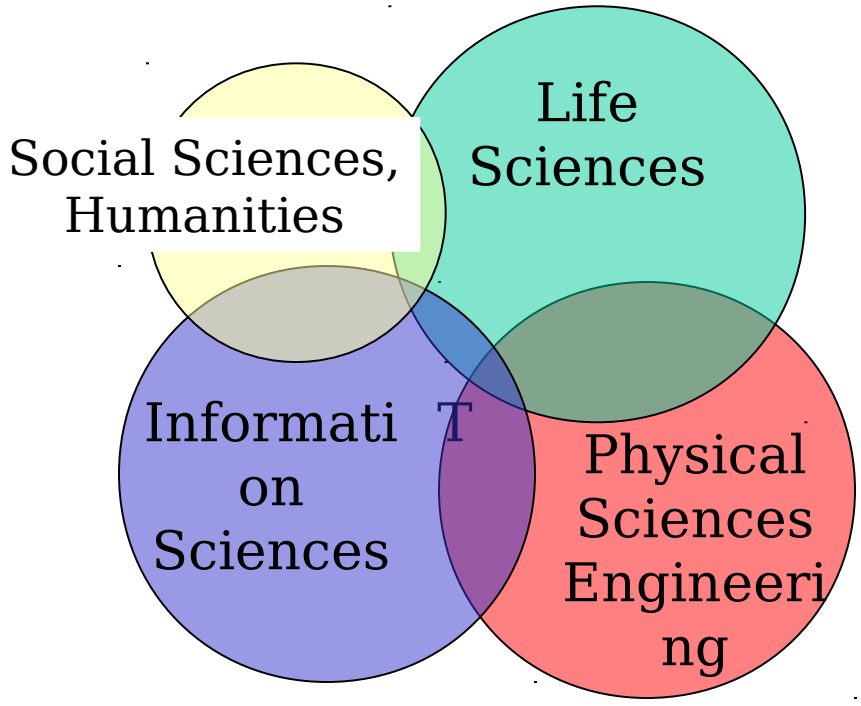
Leibnitz receives credit ...

# *Changing Scales and Contexts of Interaction and Collaboration*

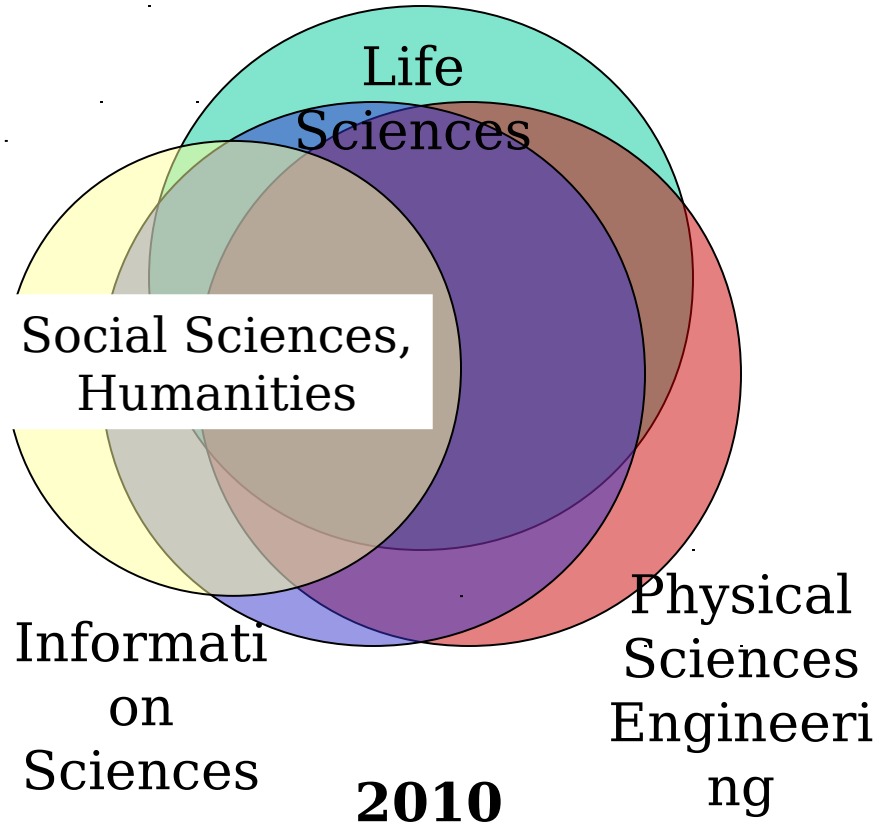


NSFNet StarTap Connections

# *A Vision of Disciplinarity: The World in 2010*

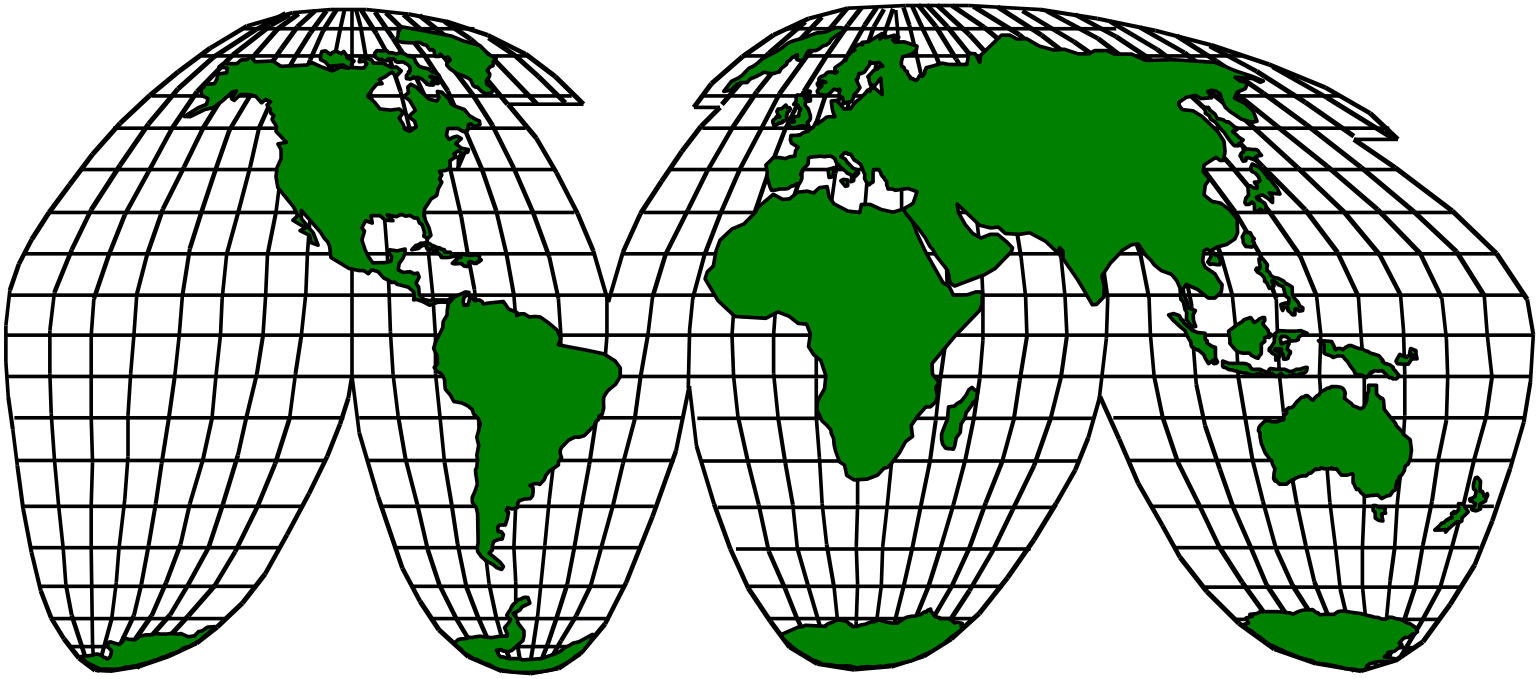



**200 0**




**2010**

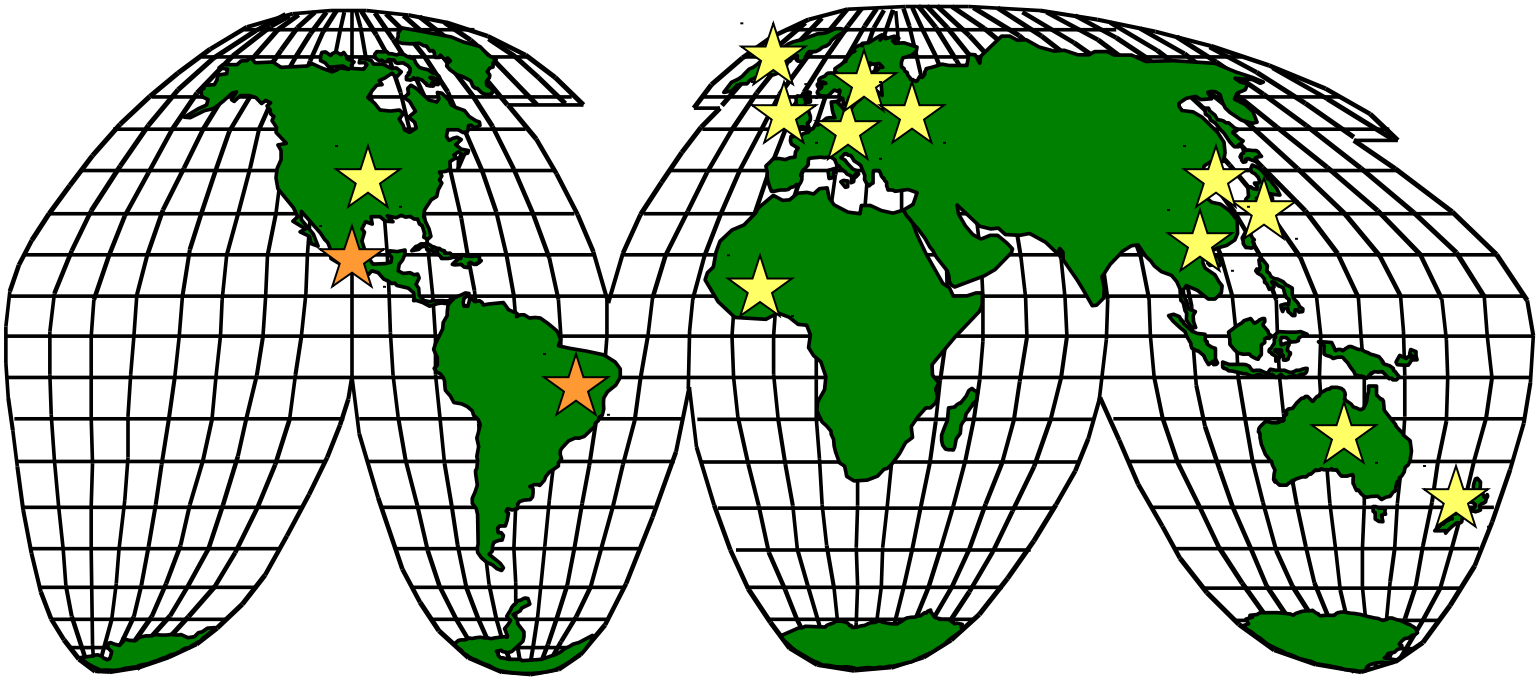
# International Digital Libraries Collaborative Research Program



**FY 1999 Competition Data**

- **~50 proposals requesting $25M**
- **~30 countries**
- **Formal Program with UK/JISC: 6 awards, $5M over 3 years**

# International Digital Libraries Collaborative Research Program



**FY 1999 Competition Data**

- **~50 proposals requesting $25M**
- **~30 countries**
- **Formal Program with UK/JISC (Circular 15/98)**

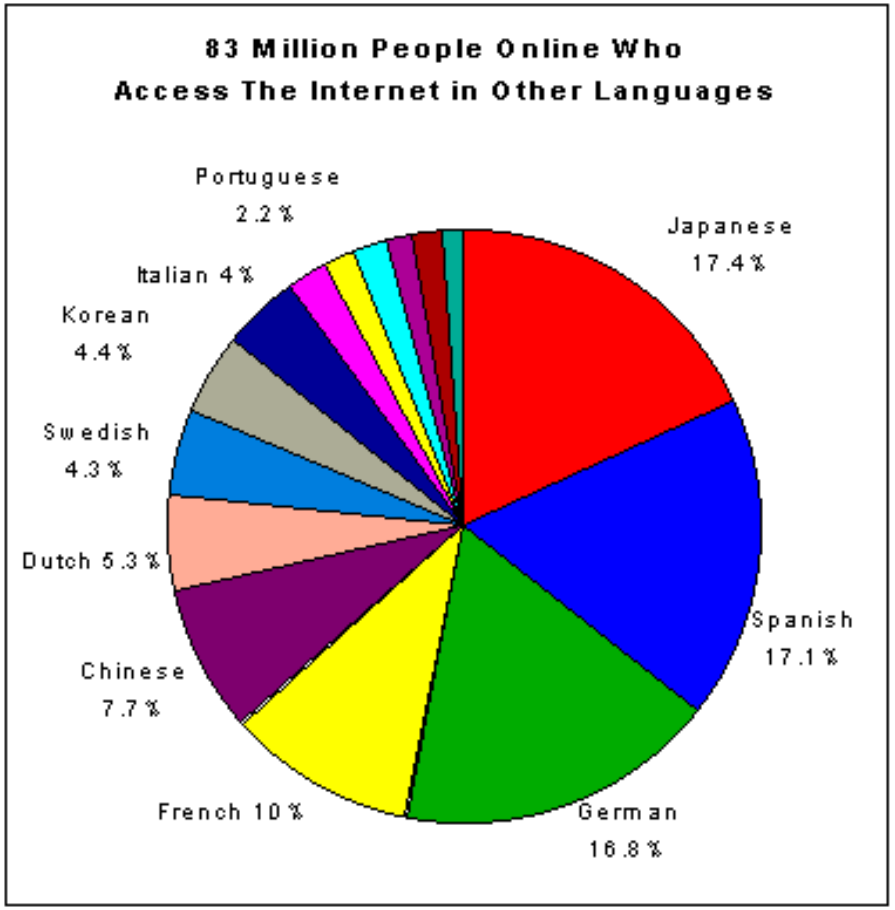
# nguages and the Internet

**April 1999**

**English 107.2M**
**56.5%**
**non-English 82.7M**
**43.5% European**
**54.9M 30.0%**

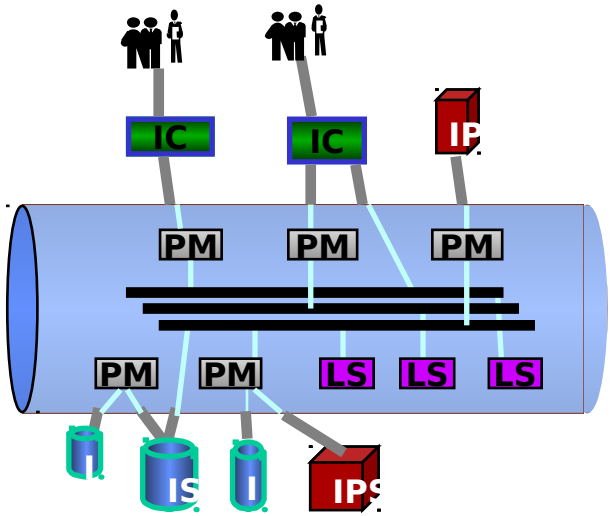
**By end of year 2000**

**English 160M**
**non-English 167M**




http://www.euromktg.com/globstats/

# *Making Digital Libraries Infrastructure Requires Dealing with heterogeneity at Many Levels**




**PM: Protocol Machine**

**LS: Library Service**

**IC: Interface Client**

**IS: Information Source**

**IPS: Information Processing Service**

**Stanford InfoBus**: CORBA distributed object technology

* objects, collections, services, platforms….

# *Making Digital Libraries Infrastructure Requires Merging Intellectual Perspectives*

**Traditional Libraries Stress:**

**Service**

**Selection, Organization, Structure for Access**

**Centralization, Standards**

**Physical objects & standard genres**

**Contemporary Technological Capabilities (e.g. WWW) Stress:**

**Flexibility, Openess**

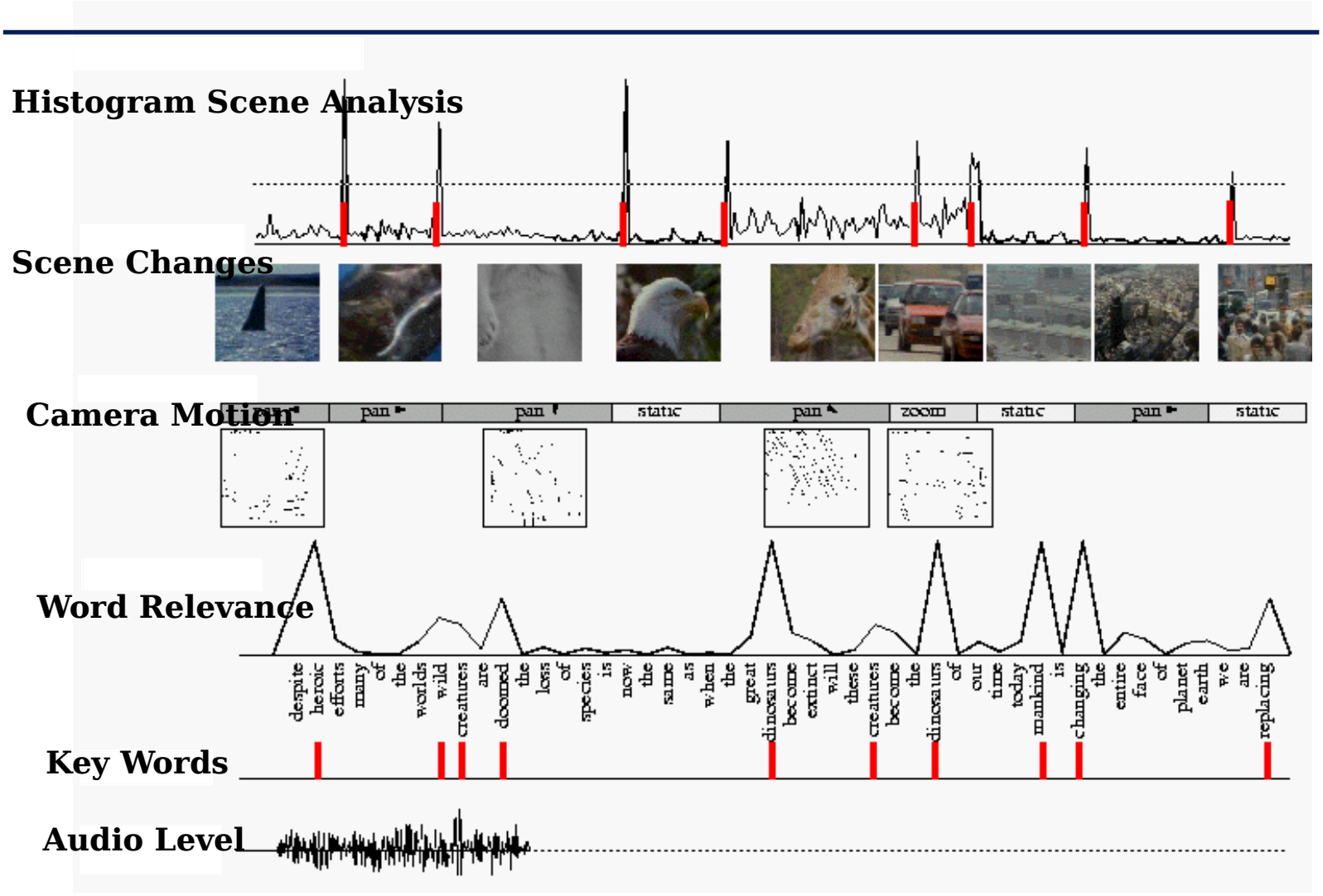
**Rapid Evolution**

**Decentralization (geographic, administrative)**

**Digital objects, old+ new genres**

**Design Space for Digital Libraries & Beyond**

# *Making Digital Libraries Infrastructure Requires Application of Integrated Technologies*


Histogram Scene Analysis
Scene Changes
Camera Motion
pan
pan
static
pan
zoom
static
pan
static
Word Relevance
despite heroic efforts many of the worlds wild creatures are doomed the loss of species is now the same as when the great dinosaurs become extinct will these creatures become the dinosaurs of our time today mankind is changing the entire face of planet earth we are replacing
Key Words
Audio Level

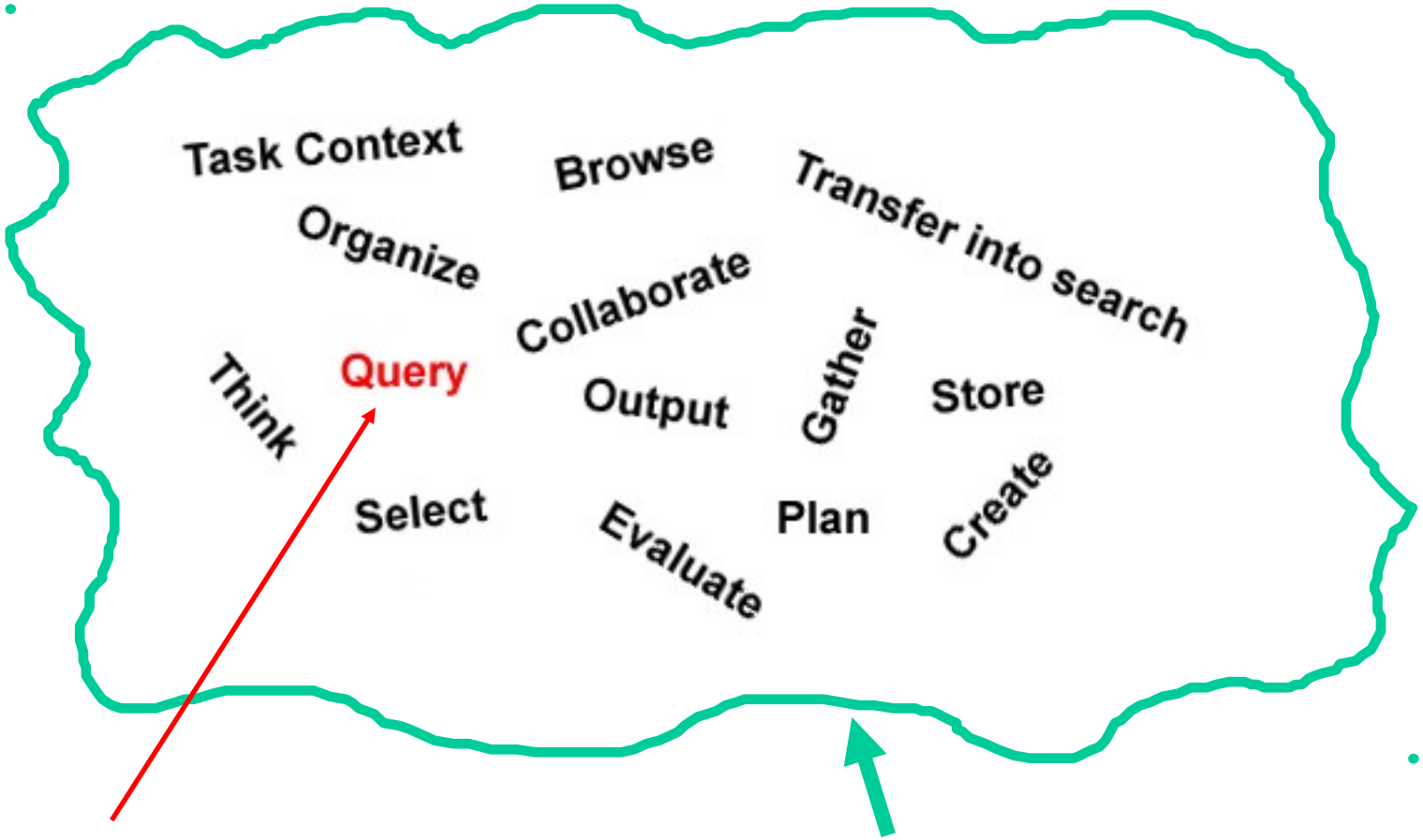
© Carnegie Mellon University 2/96

# Making Digital Libraries Infrastructure Requires Building Large Collections of Diverse Information

UC Berkeley Testbed

| Type | Examples | June 96 | | Dec 96 | |
|---|---|---|---|---|---|
| Documents | articles, EIRs, water reports | 40,900 pp. | 20 GB | 96,600 pp | 48GB |
| Images | DWR<br>wildflowers<br>Corel<br>Habitats<br>Total | 14,838<br>2,905<br>22,000<br>0<br>39,743 | 238GB | 15,506<br>7,437<br>28,101<br>158<br>52,000 | 306GB |
| Aerial photos | Suisun March Sac-SJ Delta | 5000 img | 3.4GB | 500 img | 3.4GB |
| Sensor Data | Delta fish flow | 30 days | .02MB | 30days | .02MB |
| GIS Data | dams, fish, watersheds, etc. | various | 50MB | various | 52MB |
| DOQs | SF Bay Area | 102 img | 5GB | 102 Img | 5GB |
| Digital Line Graphs | SF Bay<br>North Coast | | | | 100MB<br>100MB |
| **Total** | | | **268GB** | | **363GB** |

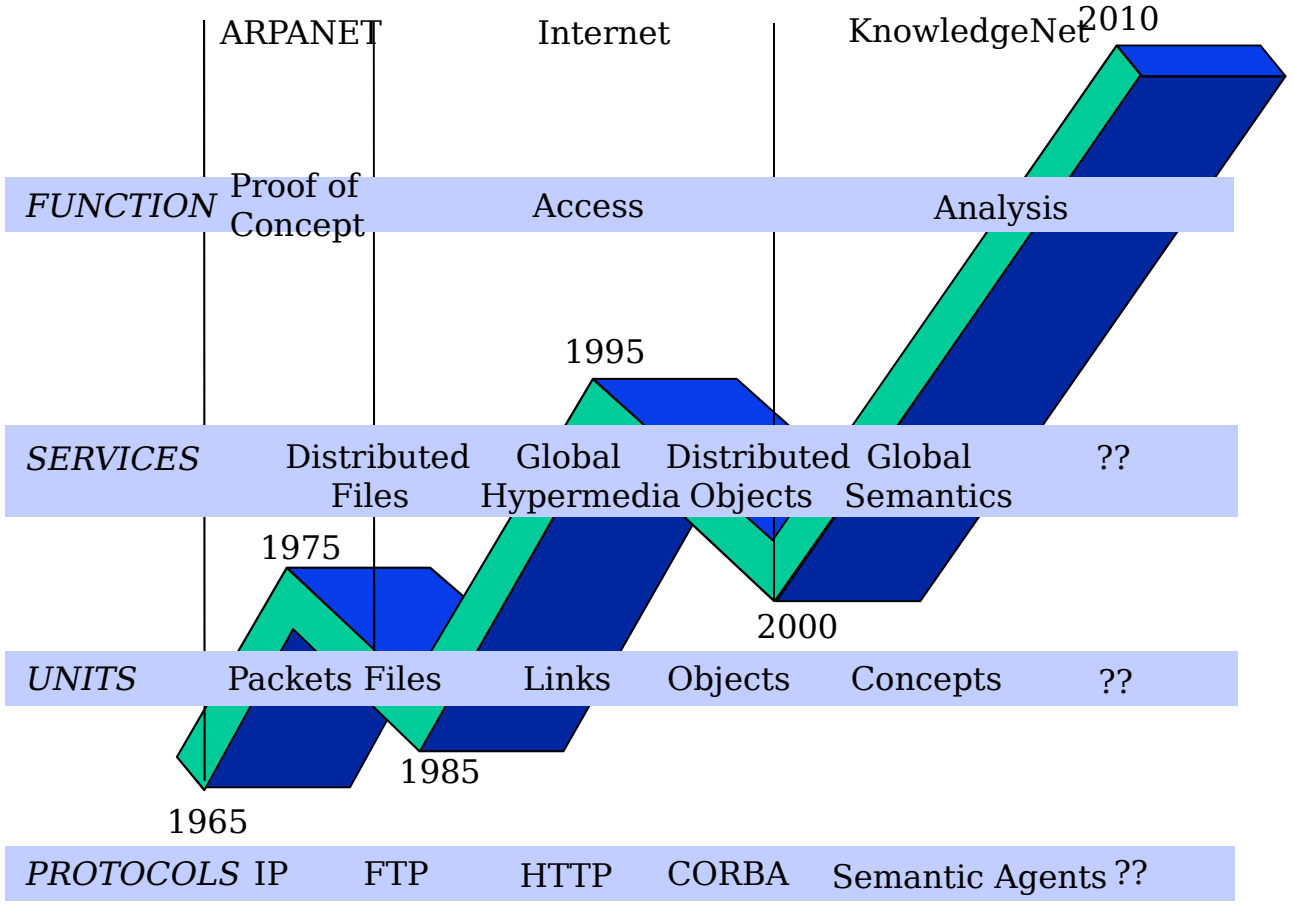
# *Making Digital Libraries Infrastructure Means Supporting More than Query*




Today's Technology Centered Systems

User and Usage Centered

# Making Digital Libraries Infrastructure Requires New Conceptualizations of the Future (imagination)



# Goals for the Future

**Gather information and build collections**

**(to better use what we have and discover what is missing…)**

**Create new communities**

**(to communicate and collaborate)**

**Make technology disappear**

**(from our awareness and experience)**

# For More Information:

**Digital Libraries Initiative National Homepage**

**http://www.dli2.nsf.gov/**

**NSF-EU International Working Groups**

**http://www.si.umich.edu/UMDL/EU_Grant/home.htm**

**http://www.iei.pi.cnr.it/DELOS//NSF/nsf.htm**

**D-Lib Magazine**

**http://www.dlib.org**

# DLI-2 Round 1 Awards

| DLI-2 Round 1 Awards | | | | |
|---|---|---|---|---|
| Award ID | PI Name | Institution | Mos. | $K |
| 9817483 | Komlduh, Mark | Mich St | 60 | 3,600 |
| 9817484 | Crane, Gregory | Tufts | 60 | 2,753 |
| 9817434 | McKeown, Kathleen | Columbia University | 60 | 5,002 |
| 9817496 | Wactlar, Howard D. | CMU | 48 | 4,000 |
| 9817432 | Smth, Terrence | UCSB | 60 | 5,400 |
| 9817799 | Garcia-Mdina, Hector | Stanford University | 60 | 4,300 |
| 9817353 | Wilensky, Robert | UC-Berkeley | 60 | 5,000 |
| 9874747 | Verba, Sidney | Harvard University | 36 | 1,800 |
| 9817416 | Lagoze, Carl | Cornell University | 48 | 2,263 |
| 9874759 | Etzioni, Oren | U Washington | 36 | 593 |
| 9817492 | Gorman, Paul | Oregon Health Sciences | 36 | 650 |
| 9817511 | Weiderhdd, Gio | Stanford University | 36 | 520 |
| 9817430 | Choudhry, Sayeed | Johns Hopkins | 36 | 530 |
| 9874771 | Armistead, Samuel G. | UC-Davis | 36 | 497 |
| 9817483 | Seales, W. Brent | U Kentucky | 36 | 500 |
| 9817444 | Bunneman, Peter | U Pennsylvania | 36 | 505 |
| 9874781 | Rowe, Timothy | U Texas, Austin | 36 | 500 |
| 9817527 | Myers, Brad | CMU | 36 | 450 |
| 9817473 | Chen, HC | U Arizona | 36 | 501 |
| 9817572 | Palakal, M. | Indiana Univ | 36 | 316 |
| 9817518 | Willer, D. | U South Carolina | 48 | 1,199 |
| Subtotal | | | | 40,894 |

*ASIS Bulletin October/November 1999*

# DLI-2 Round 1 Awards - International

| DLI-2 International Awards (FY99) | | | | |
|---|---|---|---|---|
| **Award ID** | **PI Name** | **Institution** | **Mos.** | **$K** |
| **9975164** | Larson, Ray | **UC-Berkeley** | **36** | **305** |
| **9905842** | Byrd, Donald | **U Massachusetts** | **36** | **494** |
| **9905935** | Hedstrom, Margaret | **U Michigan** | **36** | **483** |
| **9906025** | Calcari, Susan | **U Wisconsin-Madison** | **36** | **480** |
| **9907892** | Lagoze, Carl | **Cornell U/ePrint** | **36** | **292** |
| **9905955** | Lagoze, Carl | **Cornell U/ILRT** | **36** | **240** |
| **Subtotal** | | | | **2,299** |

*ASIS Bulletin October/November 1999*

# DLI-2 Round 1 Awards - Undergraduate Emphasis

| DLI-2 Undergraduate Emphasis Awards | | | | |
|---|---|---|---|---|
| **Award ID** | **PI Name** | **Institution** | **Mos.** | **$K** |
| **9817406** | Agogino, Alice | **UC-Berkeley** | **12** | **200** |
| **9816026** | Maly, Kurt | **Old Dominion Univ** | **12** | **80** |
| **9816644** | Kappelman, John | **U Texas, Austin** | **24** | **287** |
| **9816649** | Druin, Alison | **U Maryland** | **24** | **287** |
| **9980130** | Owen, Scott | **Georgia St** | **36** | **330** |
| **9980116** | Agogino, Alice | **UC-Berkeley** | **24** | **400** |
| **9979967** | Wittenberg, Kate | **Columbia U Press** | **36** | **581** |
| **9980049** | Graves, William | **College Research** | **24** | **1143** |
| **Subtotal** | | | | |

*ASIS Bulletin October/November 1999*

# LI-2 Principal Investigator Departments

| DLI-2 Principal Investigator Departments | | |
|---|---|---|
| Anthropology | Biomedical Information | Classics |
| Computer Science | Economics | English |
| Fine Arts | Geography | Geological Sciences |
| Government | Electrical Engineering | Environmental Science |
| History | Information Management | Information Studies |
| Language Technology | Library & Information Science | Linguistics |
| Management Info. Systems | Medical Informatics | Political Science |
| Psychology | Religious Studies | Robotics |
| Sociology | Spanish | Teacher Education |

*ASIS Bulletin October/November 1999*

# DLI-2 Technology Focus by Project

| | | | |
|---|---|---|---|
| 3-D Modeling | California Santa Barbara, Texas Austin | Linking | Cornell (intl – ePrint) |
| Access Control | California Berkeley | Log (Trace) Analysis | Oregon Health Sciences |
| Agents | Indiana Bloomington, Washington | Mobile Computing | Stanford |
| Archiving/ Preservation | South Carolina, U. Michigan (intl) | Multimedia Fusion | CMU, Columbia |
| Audio Retrieval | Johns Hopkins, Michigan State, U Mass Amherst (intl) | Natural Language Processing | Columbia |
| Classification, Clustering | Arizona | OCR | California Berkeley, Johns Hopkins |
| Data(Access) Services | Harvard | Parallel Processing | Arizona |
| Digital Video | CMU | Protocols | Stanford |
| Economic Models | California Berkeley, Stanford | Personalization | Columbia |
| Electronic Notebooks | California Berkeley | Provenance | Penn. |
| Federation | California Berkeley (intl), Cornell, U Wisconsin-Madison (intl) | Restoring Manuscripts | Kentucky |
| Geographic Info. Systems | California Santa Barbara | Speech Processing | California Davis, Michigan State |
| Images | California Berkeley, California Santa Barbara, Kentucky, Stanford, Texas Austin | Summarization | CMU, Columbia |
| Information Filtering | Indiana, Stanford | Text Analysis | Tufts |
| Information Visualization | CMU | Video Editing | CMU |
| Learning Contexts | California Santa Barbara Barbara | | |

*ASIS Bulletin October/November 1999*

# DLI-2 Content Focus by Project

| | |
|---|---|
| Bibliographic Records | Arizona |
| Engineering Education | California Berkeley |
| EPrints | Cornell (intl ePrint) |
| Folk Literature | California Davis |
| Geo-referenced Info. | California Santa Barabara |
| Health Care | Oregon Health Sciences |
| Humanities | Tufts, Kentucky |
| Library Reference | Washington |
| Medical Images | Stanford |
| Mixtures of Media | California Berkeley (intl), Cornell (intl ILRT) |
| Patient Records | Columbia |
| SheetMusic | JohnsHopkins, U Mass. Amherst (intl) |
| Skeletons | Texas Austin |
| Simulations | South Carolina |
| Social Science Data | Harvard |
| Speech | Michigan State |
| Video | CarnegieMellon |
| Web | Arizona, Pennsylvania, Washington |
| X-ray CT Scans | Texas Austin |

*ASIS Bulletin October/November 1999*